Sound analysis and detection, and the potential for precision livestock farming - a sheep vocalization case study

James C. Bishop[*1], Greg Falzon[1], Mark Trotter[2], Paul Kwan[1], Paul D. Meek[3]

1 University of New England, Armidale, NSW, Australia

2 University of Central Queensland, Rockhampton, QLD, Australia

3 NSW Department of Primary Industries, PO Box 530 Coffs Harbour, NSW, Australia

Abstract

Livestock vocalizations contain a wealth of information pertaining to welfare state and behaviour. Acoustic monitoring is non-invasive and has potential for numerous Precision Livestock Farming (PLF) applications. A key step in the development of a PLF acoustic monitoring system is the development of stock vocalization detection and classification algorithms. To this end, an algorithm based on Mel-Frequency Cepstral Coefficients (MFCCs) and Support Vector Machines (SVMs) was created. Audio data was acquired from a sheep farming enterprise, reflecting realistic operating conditions. Algorithm performance was across three experiments: (i) sheep vocalization classification, (ii) adult vs. juvenile classification, (iii) multi-animal vocalization. Performance in experiments (i) and (ii) was very high (>98% accuracy, stratified 10-fold cross-validation). A novel probability-based approach is proposed to handle the difficult problem of experiment (iii). The use of a threshold allows application-specific customization of class classification distribution. By use of the MFCC-SVM algorithm it is entirely possible to detect and classify sheep vocalizations in noisy environments. These results, combined with examples from the literature, show that sound analysis and detection holds promise for PLF.

Background

PLF applies process engineering principles to livestock management, utilizing senor-based monitoring inputs, automation, and intelligent systems (Wathes et al. 2008). A wide range of sensors can be used to gather information about livestock enterprises, including, but not limited to, sound (Berckmans 2014), image (Sadgrove et al. 2017), accelerometer (Barwick et al. 2016), and GPS (Falzon et al. 2013). The data steams obtained can be analyzed to find predictive or discriminatory features, which can be used to train statistical or machine learning models that provide information concerning states, conditions, or traits of interest (Wathes et al. 2008). An emerging area of PLF research is focused on the automated analysis and detection of acoustic events: sound contains an abundance of data concerning livestock and their management.

Audio monitoring and analysis has been successfully applied to a variety of PLF problems presented in the literature, providing a non-invasive means of monitoring the biological responses of livestock (Exadaktylos et al. 2014). Pig coughs contain information indicative of respiratory infection (Ferrari et al. 2008), with various systems have been developed to detect and diagnose disease (Chung et al. 2013b, Exadaktylos et al. 2011). Early recognition of respiratory infection has also been demonstrated in calves (Vandermeulen et al. 2016). Acoustic analysis has been used substantially in the determination of feed intake, particularly in cattle (Andriamasinoro et al. 2016), with much of the focus on detecting bite and chew activity (Chelotti et al. 2016, Clapman et al. 2011, Milone et al. 2012). Similar work has been conducted in free-ranging cattle, sheep, and goats (Navon et al. 2013), as well as the ingestive sounds of sheep (Milone et al. 2009). Real-time acoustic monitoring has been used to determine the feed intake (Aydin et al. 2015), and short-term feeding behaviour (Aydin and Berckmans, 2016) of broiler chickens. Reproduction-related events have also been detected in cattle, namely determining estrus via vocalization analysis (Chung et al. 2013a, Lee et al. 2014). Specific cattle vocalizations have been monitored as a means of welfare assessment, such as “murmuring” during resting and ruminating behaviour as an indicator of “good welfare” (Meen et al. 2015). Clearly, sound analysis and detection can be used for a wide array of applications in the PLF space, from giving insight into singular behaviours and welfare states, to detecting production critical events on-farm.

A test case in acoustic monitoring: ewe-lamb interactions

Livestock vocalizations possess vast amounts of information related to welfare state (Exadaktylos et al. 2011). Extensive research has been conducted into understanding sheep vocalizations, with particular attention paid to the communication between ewes and their lambs (Sèbe et al. 2010, Searby and Jouventin, 2003), and how this relates to aspects of their relationship, such as preferential nursing (Sebe et al. 2008). The majority of this work has focused on the ewe's role in communication, but recent research has delved into the lamb's contribution (Morton et al. 2017). An algorithm that is capable of detecting sheep vocalizations, with the added ability to categorizing them as either emitted by an adult or juvenile, would be highly beneficial to researchers, and be the first step in the development of a sheep vocalization monitoring system.

Methods

Experiment overview

Three experiments were conducted in order to develop and test an algorithm capable of meeting the requirements set out in Section 1.1. These were as follows:

1. Sheep vocalization classification
2. Adult vs juvenile sheep vocalization classification.
3. Multi-animal (mixed adult and juvenile) vocalization classification.

The proposed algorithm uses Mel-Frequency Cepstral Coefficients (MFCCs) (Davis and Mermelstein, 1980) combined with a Support Vector Machine (SVM) classifier (Vapnik, 1995).

Data acquisition

Field recordings were conducted at a sheep production enterprise using a Song Meter SM3 recording unit (Wildlife Acoustics, 2010), producing 720 hours of 16-bit / 16KHz audio data. The recording unit was placed in a static location, with livestock free to move and graze. Background noise was not controlled. The data collected represents a realistic example of the acoustic landscape that would be encountered during regular farm operation.

Data files were manually scanned to identify the time periods that contained the most vocalization activity of both adults and lambs. A sequential subset of 4 hours was manually processed to extract instances of vocalizations by adults, lambs, or both (i.e. multiple individuals vocalizing in close proximity). A one second window was used, as it adequately captured complete utterances, without the need to segment over multiple windows. All instances were exported and up-sampled to 16-bit / 44.1KHz quality, in WAVE format, using Audacity audio editing software (Audacity, 2017).

Data sets

Sheep vocalization classification

A vocalization class consisting of adults, juveniles, and mixed (i.e. adults and juveniles vocalizing in close proximity) was compared to a class consisting of the most commonly occurring non-vocalization sounds encountered during data extraction. The negative class is composed of birds (mixed species), noise (e.g. human activity, stock movement, digital distortion), and wind. Although it would be ideal to discriminate between ewes and lambs, the flock monitored was of mixed genders and ages, and this level of classification was not possible, largely due to the difficulty of identify vocalizing individuals when using static recording units for continuous data collection. This data set is summarized in (Table 1).

Table 1. Sheep vocalization data set summary

1: Positive	-1: Negative			Total	
Sheep	Birds	Noise	Wind	Train	Test
1655	567	687	401	2982	330

The vocalization class from Section 2.3.1 was categorized into 3 subclasses: adults, juveniles, and mixed. Subclass classification was determined aurally by the operator. The mixed class was omitted, and the remaining 2 classes were compared. This data set is reported in (Table 2).

Table 2. Adult vs juvenile data set summary

1: Positive	-1: Negative	Total	
Adults	**Juveniles**	**Train**	**Test**
720	238	863	95

Mixed vocalization classification

Using the model constructed from the data in (Table 2), all mixed vocalization instances were classified as either adult, juvenile, or unknown; this class was not used to train the model. (Table 3) outlines the data set used for the experiment. To better visual the data, Principal Component Analysis (PCA) (Wold, 1987) was used to project the data set in (Table 2) into 2 dimensions.

Table 3. Mixed vocalization data set summary

1: Positive	-1: Negative	0: Unknown	Total	
Adults	**Juveniles**	**Mixed**	**Train**	**Test**
720	238	697	863	792

Mel-frequency cepstral coefficients (MFCCs)

MFCCs (Davis and Mermelstein, 1980) were used to extract compact features from each audio instance. The steps undertaken to produce MFCCs from audio data are shown in (Fig 1), and outlined in detail in (Young et al. 2002). The parameters used to produce the MFCCs are given in (Table 4). In sound recognition problems, it is common to segment the sound into frames and apply a window function, producing MFCCs for each frame (Sharan and Moir, 2016). Through experimentation, it was found that this was unnecessary for a broad vocalization detection problem, and high accuracy was achieved without segmentation. Using this approach, feature vectors with 12 dimensions were obtained. MFCC feature extraction was implemented using Matlab software (The MathWorks, 2017).

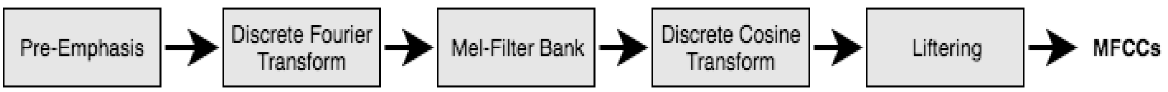

Figure 1. The steps used to produce MFCCs for a single sound window

Table 4. The parameters used when extracting MFCCs

Alpha Coef	Channels	Freq Range	Cepstral Coefs	Sine Lifter
0.97	20	300 - 6000	13	22

Support vector machines (SVMs)

SVMs (Vapnik, 1995) are widely used in sound classification tasks (Sharan and Moir, 2016). The Radial Basis Function (RBF) kernel has been shown to be a valid choice for many problems (Hsu et al. 2010). A C++ implementation of a C-SVM using an RBF kernel, utilizing the LibSVM library (Chang and Lin, 2011), was used in all experiments.

SVM Training

To train a C-SVM with an RBF kernel, 2 parameters must be selected, namely ‘C’ and ‘gamma’. A grid search was conducted using 10-fold cross-validation on the training set. The parameter combination with the highest accuracy was selected to train the model. As the adult vs. juvenile data set (Table 2) contains unbalanced data, a weight value was applied to the juvenile class to further penalize misclassification. (Table 5) displays the parameters used for each experiment.

Table 5. SVM parameters used for each experiment

Sheep Vocalizations		Adult vs Juvenile & Mixed Probability		
C	Gamma	C	Gamma	Weight (-1)
8.0	2.0	8.0	0.5	3.013953488

SVM testing

In order to adequately test the model’s generalization and performance, stratified 10-fold cross-validation (S10-CV) (Kohavi, 1995) was used during testing. In this approach, 10-fold cross-validation is undertaken for the entire data set, while maintaining the original class distribution for each fold. The average performance metrics are the calculated over all folds of the data. In order to classify mixed vocalization instances, posterior class probabilities were obtained following (Platt, 1999), with a confidence threshold applied: in this case 0.7 was used.

Results

Sheep vocalizations

Table 6. Results from performing S10-CV on the Sheep vocalization data set

	TPR	FPR	TNR	FNR	Accuracy	Precision	G-Mean	F-1 Score
M:	99.15	0.69	99.31	0.85	99.22	99.46	99.30	0.99
VAR:	0.26	0.98	0.98	0.26	0.28	0.59	0.22	0.00
STD:	0.51	0.99	0.99	0.51	0.53	0.77	0.47	0.00
CV:	0.01	1.43	0.01	0.60	0.01	0.01	0.00	0.00

Adult vs juvenile vocalizations

Table 7. Results from performing S10-CV on the adult vs juvenile vocalization data set

	TPR	FPR	TNR	FNR	Accuracy	Precision	G-Mean	F-1 Score
M:	99.03	3.91	96.09	0.97	98.32	98.77	98.90	0.99
VAR:	1.74	22.89	22.89	1.74	2.27	2.19	0.98	0.00
STD:	1.32	4.78	4.78	1.32	1.51	1.48	0.99	0.01
CV:	0.01	1.22	0.05	1.36	0.02	0.01	0.01	0.01

Mixed vocalization classification

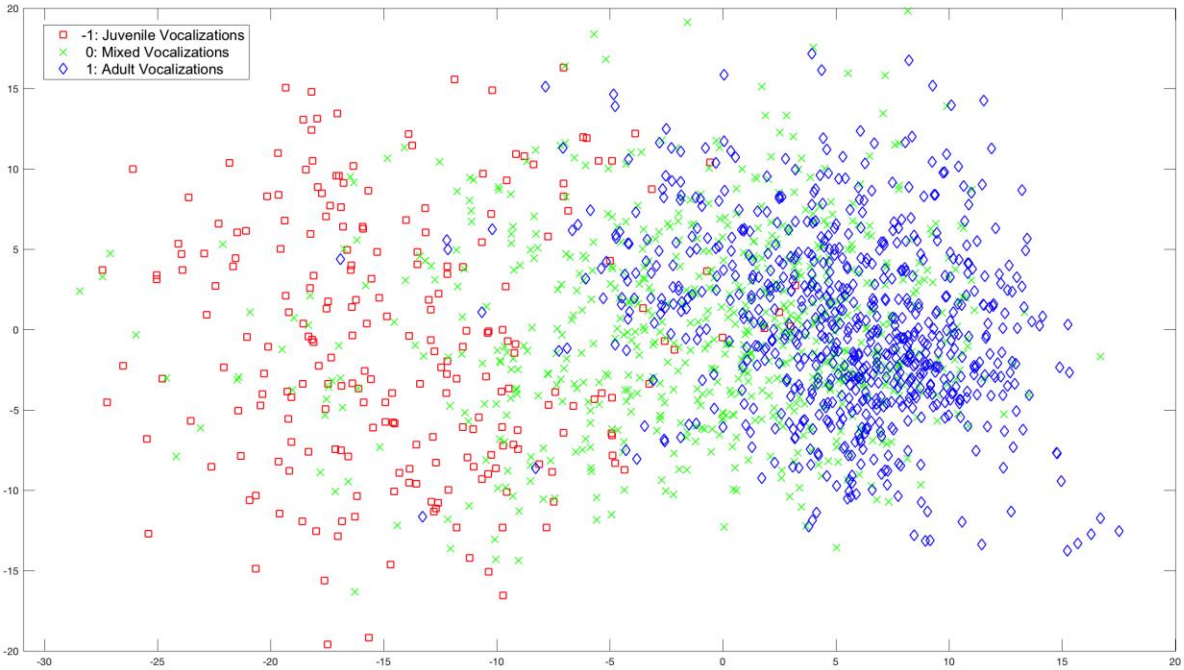

Figure 2. Results from transforming the adult, juvenile, and mixed vocalization data set into 2 dimensions using PCA

Table 8. Classification of mixed vocalization instances, using a probability threshold value of 0.7

1: Adult	-1: Lamb	0: Unknown
424	189	84

Discussion

The results shown in (Table 6) demonstrate that it is possible to accurately detect sheep vocalizations in a noisy environment. Vocalization detection performance was shown to be high is all performance metrics across 10 stratified folds of data, suggesting the model is well generalized.

(Table 7) shows that discrimination between adults and juveniles can be achieved, with high results in all metrics. Of interest is the balanced nature of accuracy, precision, geometric mean, and F-1 score.

The model displays a slightly lower ability in determining juveniles, with a corresponding higher rate of false positive classification: this is likely due to the smaller population size of juveniles in the data set. Nevertheless, performance was high, and variation was low, across all folds of data, demonstrating that the model is robust.

(Fig 2) and (Table 8) assert the challenging nature of classifying mixed vocalization instances. As mixed samples contain varying quantities of both the adult and juvenile classes, it can be viewed as a gradient between the two: some instances may contain predominantly adults, or vice versa. By applying a binary class approach to the problem, an applicable solution was found. The selection of the probability threshold will directly affect the number of instances classified to each class, and the correct setting will be entirely application-specific. By allowing manual selection, this parameter can be tuned by the user to suit their needs.

Conclusion

It has been shown that sheep vocalizations can be accurately detected and classified using data acquired from a real-world farm setting.

Future research will focus on the implementation of more generalized vocalization detection algorithms, capable of being applied to different livestock species, with a focus on low-resource systems. The algorithm will then be extended to real-time detection and segmentation applications. The problem of mixed vocalization classification will be explored by applying signal separation techniques to identified instances, with the aim of extracting individual vocalizations.

Acknowledgments

We thank Mr. D. Schneider for his assistance with data collection. Data collection was performed according to UNE Animal Ethics Committee Approval AEC15-135. Mr. J. Bishop is supported by an Australian Postgraduate Award.

References

Andriamasinoro ALH, Bindelle J, Mercatoris B, Lebeau F 2016. A review on the use of sensors to monitor cattle jaw movements and behavior when grazing. Biotechnology, Agronomy, Society and Environment 20: 273–286.

Aydin A, Bahr C, Berckmans D 2015. A real-time monitoring tool to automatically measure the feed intakes of multiple broiler chickens by sound analysis. Computers and Electronics in Agriculture 114: 1–6.

Aydin A, Berckmans D 2016. Using sound technology to automatically detect the short-term feeding behaviours of broiler chickens. Computers and Electronics in Agriculture 121: 25–31.

Barwick JD, Lamb DW, Trotter M, Dobos R, Welch M 2016. Ear deployed accelerometer behaviour detection in sheep. 13th International Conference on Precision Agriculture, Missouri, USA. ISPA.

Berckmans D 2014. Precision Livestock Farming Technologies for Welfare Management. Revue Scientifique et Technique 33: 189–196.

Chang CC, Lin CJ 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2: 1–27.

Chelotti JO, Vanrell SR, Milone DH, Utsumi SA, Galli JR, Rufiner HL, Giovanini LL 2016. A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle. Computers and Electronics in Agriculture 127: 64–75.

Chung Y, Lee J, Oh S, Park D, Chang HH, Kim S 2013a. Automatic Detection of Cow's Oestrus in Audio Surveillance System. Asian-Australas J Anim Sci 26: 1030–1037.

Chung Y, Oh S, Lee J, Park D, Chang HH, Kim S 2013b. Automatic detection and recognition of pig wasting diseases using sound data in audio surveillance systems. Sensors (Basel) 13: 12929–12942.

Clapman WM, Fedders JM, Beeman K, Neel JPS 2011. Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle. Computers and Electronics in Agriculture. 76 p.

Davis SB, Mermelstein P 1980. Comparison of parametric representations for monosyllabic sord recognition in continuously spoken sentences. IEEE Trans. on Acoustics, Speech, and Signal Processing 28: 357–366.

Exadaktylos V, Silva M, Berckmans D 2014. Automatic identification and interpretation of animal sounds: application to livestock production optimisation.

Exadaktylos V, Silva M, Ferrari S, Guarino M, Berckmans D 2011. Sound localisation in practice: an application in localisation of sick animals in commercial piggeries. Advances in Sound Localisation. InTech.

Falzon G, Schneider D, Trotter M, Lamb D 2013. A relationship between faecal egg counts and the distance travelled by sheep. Small Ruminant Research 111: 171–174.

Ferrari S, Silva M, Guarino M, Berckmans D 2008. Analysis of cough sounds for diagnosis of respiratory infections in intensive pig farming. American Society of Agricultural and Biological Engineers
51: 1051–1055.

Hsu CW, Chang CC, Lin CJ 2010. A practical guide to support vector classification. Taiwan: National Taiwan University.

Kohavi R 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. 14th International Joint Conference on Artificial Intelligence 2: 1137–1143.

Lee J, Zuo S, Chung Y, Park D, Chang HH, Kim S 2014. Formant-based acoustic features for cow's estrus detection in audio surveillance system. 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 26–29 August 2014. Pp. 236–240.

Meen GH, Schellekens MA, Slegers MHM, Leenders NLG, Van Erp-Van der Kooij E, Noldus LPJJ 2015. Sound analysis in dairy cattle vocalisation as a potential welfare monitor. Computers and Electronics in Agriculture 118: 111–115.

Milone DH, Galli JR, Cangiano CA, Rufiner HL, Laca EA 2012. Automatic recognition of ingestive sounds of cattle based on hidden Markov models. Computers and Electronics in Agriculture
87: 51–55.

Milone DH, Rufiner HL, Galli JR, Laca EA, Cangiano CA 2009. Computational method for segmentation and classification of ingestive sounds in sheep. Computers and Electronics in Agriculture 65: 228–237.

Morton CL, Hinch G, Small A 2017. Distress vocalization delay in the neonate lamb as a neurobehavioral assessment tool. Developmental Psychobiology. Pp. 1–12.

Navon S, Mizrach A, Hetzroni A, Ungar ED 2013. Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. Biosystems Engineering 114: 474–483.

Platt JC 1999. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. Advances in Large Margin Classifiers 10: 61–74.

Sadgrove EJ, Falzon G, Miron D, Lamb D 2017. Fast object detection in pastoral landscapes using a Colour Feature Extreme Learning Machine. Computers and Electronics in Agriculture 139: 204–212.

Searby A, Jouventin P 2003. Mother-lamb acoustic recognition in sheep: a frequency coding. Proceedings. Biological sciences / The Royal Society 270: 1765–1771.

Sebe F, Aubin T, Boue A, Poindron P 2008. Mother-young vocal communication and acoustic recognition promote preferential nursing in sheep. Journal of Experimental Biology 211: 3554–3562.

Sèbe F, Duboscq J, Aubin T, Ligout S, Poindron P 2010. Early vocal recognition of mother by lambs: contribution of low- and high-frequency vocalizations. Animal Behaviour 79: 1055–1066.

Sharan RV, Moir TJ 2016. An overview of applications and advancements in automatic sound recognition. Neurocomputing 200: 22–34.

The Mathworks I 2017. MATLAB R2017a [Online]. Natick, Massachusetts, United States. Available: https://au.mathworks.com/products/matlab.html [Accessed].

Vandermeulen J, Bahr C, Johnston D, EearleyB, Tullo E, Fontana I, Guarino M, Exadaktylos V, Berckmans D 2016. Early recognition of bovine respiratory disease in calves using automated continuous monitoring of cough sounds. Computers and Electronics in Agriculture 129: 15–26.

Vapnik VN 1995. The nature of statistical learning theory. New York, NY, USA.

Wathes CM, Kristensen HH, Aerts JM, Berckmans D 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Computers and Electronics in Agriculture 64: 2–10.

Wildlife acoustics I 2010. Song Meter SM2 [Online].
Available: http://media.nhbs.com/equipment/sm2-manual.pdf [Accessed].

Wold S 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems
2: 37–52.